DOCUMENT RESUME

ED 212 653 TM 820 040

AUTHOR Wanous, Donna S.; Mehrens, William A.
TITLE Helping Teachers Use Information: The Data Box Approach.
INSTITUTION National Council on Measurement in Education, Washington, D.C.
PUB DATE 81
NOTE 12p.
AVAILABLE FROM National Council on Measurement in Education, USPS823120, 1230 17th Street N.W., Washington, DC 20036. (Subscription: $6.00 per year; $2.50 single issue).
JOURNAL CIT National Council on Measurement in Education; v12 n4 Win 1981

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Classroom Techniques; *Delivery Systems; *Educational Assessment; Elementary Education; *Elementary School Teachers; Evaluative Thinking; *Information Utilization; Instructional Improvement; *Instructional Materials; Measurement Objectives; Teacher Education; Testing
IDENTIFIERS *Data Box

ABSTRACT

Describing the findings and insights gained from a two-year research and development project entitled "Integrating Assessment with Instruction," this document focuses on the current measurement needs of teachers and the instructional processes for meeting those needs. The article is divided into four segments: (1) a description of the literature in the field of educational measurement that guided the development of the project; (2) a description of the project, its goals and components; (3) a discussion of the results from the field trials; and (4) a set of reflections about recommendations for the professional development of teachers in the content area of educational measurement. The "Data Box" is introduced and described as an instructional package having teachers investigate the use of assessment data in a variety of instructional decision-making situations. There are six components: a taxonomy, a set of vigrettes, a document file, a memo pad, an interpretive manual, and an applications manual. Major measurement concepts teachers need to know, and a thorough evaluation plan, are given. (Author/CE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
the Editor
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED212653

NCME
measurement in education
A SERIES OF SPECIAL REPORTS OF THE NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION

Volume 12, No. 4 Winter, 1981

Helping Teachers Use Information: The Data Box Approach

Donna S. Wanous and William A. Mehrens
Michigan State University

ABOUT THIS ISSUE

Donna S. Wanous

William A. Mehrens

In the beginning there was teaching, and then there was testing. But somehow the two have rarely been linked in any meaningful and sustained way. This, in turn, has given rise to a concern for "over-testing" and a suspicion of standardized tests. Teachers have used the results of standardized tests for almost sixty years but all too often have tended to overgeneralize from inadequate data, and have — at times — not thoroughly understood the standard scores they were called upon to interpret.

Professional development has kept pace with classroom instruction, but not with the skills needed in the use of test results for classroom decision-making. All too often when tests become the agenda of an inservice program the experience is a hortatory one filled with charismatic speakers attempting to convince teachers of their need for diagnostic-prescriptive teaching coupled with the use of test results. But as with all good sermons from the pulpit, the feeling of enthusiasm wanes in direct proportion to the distance from the lectern.

Drs. Wanous and Mehrens recount for us a unique research and uevelopment program in which they have participated for the past two years. The development of *The Data Box* has resulted in the production of a comprehensive survey of test use literature that covered the past sixty years of research in United States schools, the conceptualization of a taxonomy which links basic measurement concepts with eight classroom contexts, and with a simulated set of materials designed to help teachers become aware of the power of data in their daily teaching. An extensive evaluation of *The Data Box* was conducted in the U.S. Department of Defense Dependents Schools. The prototype tested in these schools is undergoing some revisions in light of this evaluation and will soon be distributed by a major test publisher.

Dr. Donna Wanous is an Assistant Professor of Teacher Education at Michigan State University and is a member of MSU's Institute for Research in Teaching. Dr. William A. Mehrens is a Professor of Counseling, Educational Psychology and Measurement, Michigan State University. This paper represents Dr. Wanous' first appearance in *ME*, and Dr. Mehrens' third paper. They are both excellent contributors and are worthy, indeed, for ending the tenure of *ME* on an high note.

HCR

TM 820 010

The purpose of this article is to describe the findings and insights gained from a two-year research and development project entitled, "Integrating Assessment with Instruction." The focus will be on those findings related to the current measurement needs of teachers and the instructional processes (delivery system) for meeting those needs.

The article is divided into the following sections: a description of the literature in the field of educational measurement that guided the development of the project; a description of the project, its goals, and components; a discussion of the results from the field trials; and a set of reflections about recommendations for the professional development of teachers in the content area of educational measurement.

What Do We Know?

In reflecting upon the key issues in the arena of educational tests and measurements during the decade of 1970's, it is clear that a major controversy surrounded standardized tests (Stetz & Beck, 1981). The war on testing comes primarily from special interest groups rather than the general public who is generally more for testing than against testing (Lerner, 1980). Leaders of the two major teachers unions take quite different positions on standardized tests. Albert Shanker, President of the American Federation of Teachers, strongly supports tests (Shanker, 1980), while the National Education Association has proposed a moratorium on standardized testing. Teachers, as a group, are much closer to the AFT position. Stetz and Beck (1981) report that only 16% of a national sample of teachers agreed with the call for a moratorium.

While this controversy has raged on among the various constituent groups, teachers (who, as Stetz and Beck point out, were rarely directly consulted in the controversy) have had to continue making instructional decisions on a daily basis. Often these decisions were being made in classrooms where the composition of the student body was rapidly shifting due to other education decisions such as mainstreaming and busing.

For many teachers, the controversy over standardized testing often removed the one potential source of assessment data with which teachers were already familiar. Many teachers reported that they had no formal standardized data to assist in their instructional decision making for students who had educational needs with which these teachers were not as familiar or trained to assess.

Thus teachers were often left to develop their own sources of assessment data upon which to make their instructional decisions. However, as several authors have stated, this task of measurement development is not one for which teachers have received a lot of adequate training at either the preservice or inservice level (Bradley, 1978, Fleming, 1979, Nicholson, Joyce, Parker & Waterman, 1976; Olejnik, 1979; and Yeh, 1978). The lack of training emphasis exists even though measurement and educational experts agree that it should be a crucial element in the education of teachers and prospective teachers.

As an example of the perceived importance of this training, school administrators and teachers at a 1978 conference sponsored by the National Institute of Education's Office of Testing, Assessment and Evaluation "emphasized the need for teacher education and inservice training on test development, selections and

interpretation" (*Development of Materials*, 1979, p. 3). This type of training was also seen as crucial to "educational change and as a catalyst for improving educational practice." (p. 4). In a study conducted by Yeh (1978) teachers reported that they wanted more training on educational tests and measurement, especially on the topic of criterion-referenced testing.

Therefore, while the need of teachers for instruction in educational tests and measurement has been established, it appears that this need has not been met at either the preservice or inservice level. In a survey of 551 preservice teacher education programs, it was reported that required courses in "testing and measurement" ranked 14 out of 21 possible topics in terms of the mean number of required credit hours (Nicholson et al., 1976).

In a paper identifying classroom measurement needs, Fleming (1979) pointed out that while the training has been neglected, the need for these measurement competencies has remained critical. She speculated that this problem may be due to the lack of relevancy of the training to the classroom situation. She stated:

> Given that preservice training of teachers is guilty of several omissions in not demonstrating the utility of evaluation in the instructional cycle, in failing to develop a content for use of measurement as an appropriate tool in instructional decision-making and in failing to utilize approaches relevant to the classroom tasks in measurement, inservice training offered to teachers in the classroom has been just as ineffective. Experience is no guarantee that teachers will develop such competencies, particularly in the area of their classroom tests (p. 4).

Yeh (1978) came to a similar conclusion:

> Another source of inertia related to the use of test results may be teachers' lack of knowledge about the ways test design and test interpretation can influence their own decision-making ... These inadequacies in teacher training are not surprising given the fact that those involved in designing and conducting that enterprise rarely have strong backgrounds in the field of measurement themselves.

The Data Box Project

Goals

In an attempt to address this apparent gap between teachers' educational measurement needs and the training they have received, the National Institute of Education and the Department of Defense Dependents Schools funded a two-year research and development project to examine the relationship between assessment and instruction. The ultimate goal of this project was to increase teachers' use of assessment data in the classroom by focusing on various methods for integrating assessment data into classroom instructional decisions.

To accomplish this goal, the following four subgoals were established:

1. to determine the current degree to which assessment data are integrated into instructional decision-making in the classroom (a review of the literature);
2. to develop a set of materials and two accompanying delivery systems that would increase teachers' knowledge and use of assessment data (*The Data Box* and Delivery Systems I and II);
3. to analyze the impact of this training system on classroom instruction; and
4. to understand what teachers define as assessment data.

It should be noted that the term "assessment data" was broadly defined in this project to mean any "systematic observation of behavior which may be collected through tests, questionnaires, observation and other techniques" (*Development of Materials*, 1979, p. 2). Thus the term assessment data refers to more than just standardized tests.

Components

Review of related research. The first step in determining how teachers use assessment information in the classroom was a review of the related research (Rudman, Kelly, Wanous, Mehrens, Clark, & Porter, 1980). This review covered 94 studies from 1922 to 1980. The review was organized around the following four questions:

1. How do teachers use test information for instructional decisions?
2. What has research indicated about the linkage between testing and teaching?
3. What is known about the successful practice of inservice training of teachers?
4. What is and has been available to teachers to aid them in interpreting and using assessment information for instructional decision making?

In general this comprehensive review revealed a literature that is uneven in quality, in scope of serious study, filled with inconsistencies, and difficult to find in the traditional channels of research. Some examples of the answers to the four organizing questions were as follows:

1. It would appear that there is a gap between what teachers think about testing and what others write about what they think.
2. Teachers' estimates of pupil performance on tests are remarkably stable and consistent. They are not nearly as prone to set expectations as earlier characterized.
3. In contrast to teachers' perceptions of their students' test scores, there is some evidence that teachers' reporting of their students' classroom interpersonal behavior is neither stable nor accurate.
4. There is conflicting evidence concerning teachers' sensitivity to the information they use in instruction. Some researchers report that teachers are sensitive to the reliability of the information they receive and can adjust their judgments of students, accordingly, while others report that teachers ignore the quality of the data.
5. There is considerable reason to believe that teachers have limited knowledge about fundamental measurement concepts.
6. Most inservice experiences are similar in

design. They consist of lectures, self-study modules and pamphlets, and are usually structured for group participation and study.

The Data Box. This instructional package is an integrated set of materials designed to have teachers investigate the use of assessment data in a variety of instructional decision-making situations. There are six components: a taxonomy, a set of vignettes, a document file, a memo pad, an interpretive manual, and an applications manual. The components revolve around the vignettes which are based on everyday classroom events.

A set of seven assumptions provided the framework for the development of *The Data Box*. The assumptions were the following:

1. There are critical incidents in the daily practice of teachers which require specific assessment information.

2. There are fundamental measurement concepts and skills that teachers need in order to accurately gather assessment information and interpret it for these critical incidents.

3. The materials contained in *The Data Box* had to be realistic and relevant to teachers' everday classroom experiences.

4. The vignettes within *The Data Box* needed to cover a wide range of classroom events in order to appeal to a cross section of teachers with different levels of skills and knowledge.

5. *The Data Box* needed to be a self-contained unit; a unit which could be disseminated with or without any accompany on-site professional development.

6. Use of *The Data Box* would require teachers to integrate all the materials with their own knowledge in resolving the vignettes.

	Relating Educational Decisions to Data	Knowing What, Where & How to Use Data in Dec. Making	Knowing the Sources of Data	Knowing Data Collection Processes	Understanding Data Characteristics	Interaction of Data With Sources and Processes	Specifics of Data Use	Knowing How to Disseminate Information	Disseminating Information to Various Groups		
CONTEXTS	CONCEPTS/VIGNETTES										
1.0 BEGINNING OF YEAR	1.1 1.1	1.2	1.1 1.2	1.2	1.1	1.2	1.1 1.2				
2.0 CLASSIFYING	2.1	2.1	2.1	2.1		2.1	2.1				
3.0 DIAGNOSIS	3.1 3.2	3.1 3.2	3.2	3.1	3.2	3.1 3.2	3.1				
4.0 PACING	4.1	4.1	4.1	4.1	4.1	4.1	4.1				
5.0 RETENTION AND PROMOTION	5.1	5.1	5.1	5.1	5.1	5.1	5.1				
6.0 EVALUATING	6.1 6.2	6.1 6.2	6.1	6.1	6.1 6.2	6.2	6.1 6.2	6.1			
7.0 PARENT CONFERENCE	7.1 7.2	7.1 7.2	7.1 7.2	7.1 7.2	7.1	7.1 7.2	7.1 7.2	7.1 7.2	7.1 7.2		
8.0 CURRICULUM PLANNING	8.1 8.2	8.1 8.2	8.1 8.2	8.2	8.1 8.2	8.1 8.2	8.1 8.2	8.1	8.1 8.2		

Figure 1. A Contextual Events/Measurement Taxonomy

2.1 Classifying/Grouping for Mathematics Instruction

The first marking period is just completed. You have turned in your report cards and are beginning to plan for subsequent instruction. During the second nine weeks you will have a student teacher. You have met with this person and, since she is very interested in the teaching of arithmetic, you have jointly decided to do some intraclass grouping in mathematics.

Given the 4th Grade Class List and all the information provided in *The Data Box* determine (1) how many groups you should set up, (2) who should be in which groups, and (3) what the instruction should be in those groups. Present a rationale for your answers.

In doing the above tasks what information in *The Data Box* did you use? What information not in *The Data Box* would have been helpful in determining your answers to the above questions?

In responding to this vignette assume that the 8 quizzes covered the following topics (see *The Data Box* for descriptions).

Quiz		Topic
1	Unit 1:	Numeration Parts A and B
2	Unit 1:	Numeration Parts C and D
3	Unit 2:	Addition & Subtraction Parts A, B and C of Addition
4	Unit 2:	Addition & Subtraction Part D of Addition, Parts A & B of Subtraction
5	Unit 2:	Addition & Subtraction Parts C, D, and E of Subtraction
6	Unit 3:	Measurement Parts A, B, and C
7	Unit 3:	Measurement Parts D, E, and F
8	Unit 3:	Measurement Parts G and H

After you have completed the tasks specified above, turn to page 11 in *The Interpretive Manual* and compare your responses to the sample responses provided there.

Figure 2. A Sample Vignette

7. Within *The Data Box* there needed to be a provision for the addition of site-specific instructional situations.

The first component, the taxonomy, is the organizational guide for *The Data Box*. The taxonomy contains the two dimensions of contexts and concepts. The contexts represent major instructional decisions that occur during the school year while the concepts refer to the measurement concepts embedded in the contexts. The cells within the taxonomy contain the vignettes (by number) that represent the intersection of the two dimensions (see Figure 1).

In each of the 13 vignettes, a classroom situation which requires an instructional decision is described. The reader is then asked to specify what he/she would do in that situation. All the vignettes and supporting materials are based on a simulated fourth-grade classroom. However vignettes are general enough to be applicable across several grade levels (see Figure 2).

In the process of arriving at a solution to the vignettes, the reader is asked to consider a set of documents. These documents consist of both formal and informal assessment data that might be available in the typical school. For example, these data include standardized achievement test results, grade book pages, reading diagnostic test results, court orders, and various memoranda regarding parental conversations.

Each *Data Box* contains a memo pad on which the readers are asked to record the date, the vignette with which they are working, the documents they selected, and their solution to the problem posed in the vignette. The readers are asked to record these data for each vignette they complete.

The interpretive manual contains the "solutions" to the vignettes. These solutions were developed by the authors of *The Data Box* as discussion generators. The solutions are not intended to be "the one right answer" but rather a model against which each user can compare his or her response to the problem posed.

The sixth component of *The Data Box* is the applications manual. This manual supplies the reader with a perspective of the system, some summary information about what is known about the use of assessment data in schools, the various ways *The Data Box* can be used, a bibliography, and an extensive glossary of measurement concepts and terms.

One of the crucial stages in the development of *The Data Box* was the selection of the measurement concepts to be included within the vignettes. It was difficult to decide just what measurement concepts to cover in

professional development. There are a lot of ways to slice the pie. Measurement specialists often think they have a pretty good idea of what teachers should know and what they do know. The difference is what they need to be taught. Teachers, however, also have ideas about what it is they know and what they need to know and their perceived needs are not always the same as the measurement specialists' perceived needs for them. It is not at all obvious which group is correct. Most likely they are both right and wrong in both camps.

Even if there were agreement about the teachers' needs, the question of what to cover in any given professional development approach would not be answered for surely the needs would be greater than could be met. Priorities would have to be set. Further some needs, while perhaps more important, are not as amenable to being met via a given professional development approach.

Our task was to choose measurement concepts that were (1) truly important, (2) perceived as important by teachers, and (3) susceptible to instruction via our chosen approach to professional development.

The initial step in determining what fundamental concepts to stress was to look at the content of a comprehensive text book (Mehrens & Lehmann, 1978). Each chapter in this text contains a set of objectives and there are a total of 213 objectives. These objectives were organized under nine major concepts that were thought to represent the underlying structure of the body of knowledge, skills, and attitudes in measurement and evaluation most germane to educators. The nine major measurement concepts are listed in Table 1.

Following this facilitative organizational task, the staff reduced the 213 objectives to a sub-set of 61. These were sent out to a variety of reviewers. Following some limited feedback this set of 61 was reduced to 31 objectives. Some examples of these 31 are as follows.

1. Recognize that measurement and evaluation are essential to sound decision making.
2. Understand the advantages and limitation of observations.
3. Appreciate the importance of reliability of data used in decsion making.
4. Construct a test blueprint.
5. Understand the instructional value of feedback.
6. Distinguish between norms and standards.

We did not include more esoteric objectives involving such skills as computations of standard errors in our final list (although that example did survive our first cut).

Delivery systems. Two very different delivery systems were developed for the implementation of *The Data Box*. Delivery system I consisted of only a set of written directions to accompany *The Data Box*. This system required that practitioners in a particular school site initiate any professional development experiences that might be provided with *The Data Box*.

The development of delivery system II contained a far more elaborate professional development component which was based on the staff's experiences with professional development, advice from an External Advisory Panel of practitioners, and the results of reviews of

Table 1. Major Measurement Concepts Teachers Need to Know

I. Recognize that educational decisions should be based on data.

II. Understand that teachers need to know what data to gather, how (where) to gather it, how to use it in their own decision making, and how to disseminate the information to other decision makers.

III. Know about various *sources* of data such as the following:
1. parents
2. previous teachers
3. support staff
4. physicians
5. student
6. current teacher himself/herself

IV. Know about *data collection processes* including the following:
1. previous records (e.g. CA. 60)
2. interview with above sources
3. observations (formal & informal)
4. testing (teacher & standardized)

V. Understand *data characteristics*
1. Relevance (a. content validity, b. predictive validity)
2. Consistency
   a) types (causes) of random errors
   b) standard error of measurement and band interpretation
3. Costs: Financial & Psychological includes ease of gathering, invasion of privacy, etc.

VI. Interaction of Data Characteristics with Source & Process
1. Know how to build instruments with appropriate data characteristics.
2. Know how to choose instruments with appropriate data characteristics.
3. Know how to analyze instruments with respect to appropriate data characteristics.

VII. Know how to use data in Decision Making
1. Know how to combine (weight) data — Statistical advantage over clinical
2. Know how to convert data to common scale
3. Understand how to interpret various types of scores such as percentiles, T scores, stanines, grade equivalents, and deviation IQ scores.
4. Understand concepts of two types of errors and their relative costs.
5. Recognize that there is no reason to diagnose unless it leads to differential treatment.

VIII. Know how to disseminate information
1. to children
2. to parents
3. to other educators
4. to the general public

IX. Understand what information should be disseminated to the above named groups.

literature that were related to successful professional development (Cruickshank, Lorish & Thompson, 1979). Based on these three sources of information, a set of 10 guiding principles was developed. Professional development experiences are more likely to be successful if they are:

1. designed for long-term involvement rather than single-shot sessions.
2. designed collaboratively by representatives of all role groups who will be participating.
3. provided at the school site rather than at a university.
4. designed to incorporate relevant research findings into the daily instructional practice of the classroom.
5. designed to receive full administrative support.
6. designed to allow for one-to-one interactions among the specialists and participants.
7. predicated on a follow-up program which utilizes local staff to continue its momentum in the classroom.
8. designed to provide an opportunity for teachers to interact with other teachers as specialists and colleagues.
9. designed to meet the needs of the individual teacher in his/her specific setting.
10. based on a plan that utilizes demonstration, supervised trials and feedback.

Each of these principles was integrated into the development of delivery system II. As a result of this integration, the original plan for the orientation work in the schools was as follows:

1. to spend two days at each school site to learn about the local situation and to establish credibility with teachers,
2. to present one-day workshops on the content and methods of using *The Data Box* to separate groups of administrators and teachers,
3. to enlist volunteer teachers who were willing to work through the set of vignettes over the succeeding three-month period,
4. to assist these volunteer teachers in beginning their work with *The Data Box*,
5. to establish a site coordinator from within the staff of each school,
6. to provide follow-up activities, and
7. to gather some baseline data from the participants.

Both delivery systems were field tested with American teachers in American schools in the DODDS system in Europe during January, 1981 to May, 1981. Unfortunately delivery system I was never fully implemented. It is difficult to determine whether this was due to the delivery system itself or merely a site specific phenomena. The literature about successful professional development would suggest the former explanation.

Delivery system II was implemented in four elementary schools and one junior high. A total of 37 teachers ultimately volunteered to participate in the three-month portion of this professional development model.

What Did We Learn?

Evaluation Plan

The evaluation component was designed to answer the following primary questions:

1. What were the reactions of teachers and other school personnel to the format and content of the materials and the delivery system?
2. After having experienced training through the use of materials with a delivery system, what was the state of teacher knowledge about measurement concepts and their application in a classroom setting?
3. As a result of training, what changes were evidenced in teacher classroom practices?

Some additional secondary questions included the following:

4. What information did the teachers select from the documents file in making their decisions?
5. How did this information differ from that given in the suggested solutions?
6. How did the teachers define assessment data?

In order to answer the primary and secondary questions, several types of evaluation data were gathered during the implementation phase of *The Data Box*. The following data were collected on a pre/post basis:

1. a 20-item test measuring knowledge of measurement concepts;
2. a one-page-evaluation form designed to assess the teachers' reactions to the delivery system and *The Data Box;* and
3. an interview survey designed to determine how teachers define assessement; how they use assessment data; and to further evaluate *The Data Box* and its delivery system.

In addition the following data were collected on a post basis only:

1. a 49-item questionnaire assessing the teachers' reactions to the specific components of *The Data Box*, and
2. the sheets from the memo pad that were completed as the teachers worked through the vignettes.

These data were analyzed using a variety of techniques ranging from analyses of variance to the use of triangulation in the interview analyses. Since in most cases several different analyses were used to answer each evaluation question, it was necessary to aggregate the data at the primary and secondary question level.

Since this was a developmental project, not a research experiment, the team realized that the results would be tentative at best. However, the staff did believe that by collecting a wide variety of relevant data and integrating the analyses across situations and teachers, some meaningful results would be obtained.

Findings

The findings are organized around each of the primary and secondary questions. Given the extensive amount of data, only summaries of the data are presented (See Rudman et al., 1981 for a more complete reporting of the data).

The teachers' reactions to *The Data Box* and delivery system II were generally positive. The teachers reported that they found the vignettes and the solutions to be realistic in reflecting classroom situations. They also stated that they liked having an extensive period of time to work on the projects with their colleagues.

The pre/post shifts in the evaluations did, however, differ across school sites. While two of the sites reported a higher but not significant positive shift, the other two sites reported a significantly negative shift. It is interesting to note that the two sites with the negative shifts had the highest ratings on the pre evaluation. While many site specific factors may have contributed to the shift, perhaps the expectations were too high in the beginning.

As a result of these data and suggestions made by the teachers, revisions of *The Data Box* are now in progress. The majority of these revisions focus on organizational factors.

Regarding changes in the state of teacher's knowledge of measurement concepts, a significant pre/post change did not result. The mean on the pretest was 11.05 and the mean on the post test was 11.46. As one of the participating teachers pointed out, perhaps the level of specificity of knowledge called for in the test was too high. She stated that while teachers in her group had increased their general measurement knowledge, that fact would not be reflected in the test.

The teachers reported several changes in their classroom practices. These changes were reported regardless of how the teachers evaluated *The Data Box* experience. Even those teachers who felt negatively about *The Data Box*, still reported positive shifts in their classroom practices and their attitudes toward assessment data.

Fifty two per cent of the teachers reported changes in their teaching (37% reported no changes and 11% were not sure). These changes included becoming more data oriented, developing their own assessment measures, increasing the use of multiple pieces of data, developing their own self assessment measures, becoming more careful in the selection of curriculum packages that include assessment measures, and using more data to make instructional decisions.

The answers to the three secondary questions were as follows:

1. When working through the vignettes, teachers generally selected the same information from the document file as the staff did. They often used more documents than were suggested that the staff use.
2. Eighty-one per cent of the teachers reported that they either agreed or mostly agreed with the given solutions to the vignettes.
3. Teachers had a very broad definition of assessment data. It ranged from how a child held a pen to sociograms to standardized tests. The definition of assessment data broadened from the beginning of the project to the end of the project. The majority of individual definitions included some form of observational data.

Using Information: Some Reflections

As the team reflected upon its experiences with the project and the data collected from the teachers, several recommendations and concerns were discussed. Several of these recommendations and concerns are offered here with respect to the content and procedures of professional development programs in educational measurement.

1. Measurement experts need to consider exactly how detailed teachers' knowledge of measurement concepts must be. Teachers continually reported that they wanted to have some guiding principles and pertinent references. Such activities as that of memorizing a formula for reliability would not be as important as understanding the concept itself and how it relates to classroom assessment of all kinds. Most professional development programs probably err on the side of presenting too many measurement concepts at too abstract and deep a level.
2. Teachers expressed a need for more knowledge about existing methods for measuring the affective dimensions of their classrooms and instruction. This would include instruction on how to construct these kinds of instruments.
3. Teachers reported being quite comfortable with their knowledge of assessment data in the curricular areas of reading and mathematics but not in other areas. In these other areas teachers stated that they depended on the assessment measures that accompanied the curricular packages they used. As a result teachers wanted more knowledge about how to evaluate the quality of these materials and how to construct their own assessment instruments.
4. How a teacher might combine a variety of assessment data from various sources was another area of need that teachers identified. How do you combine standardized test score data with observational data in a meaningful way when making an instructional decision?
5. Two different kinds of knowledge were delineated as the teachers worked with *The Data Box*. The first one was knowledge about how to read the forms that accompany standardized materials. For example, this would include knowing where to look on the form for a particular type of score, or how to find an individual student's scores versus classroom scores. The second kind of knowledge is knowing what these various terms, e.g., percentile scores, on the form mean. Teachers recognize the need for professional development programs that cover both types of knowledge.
6. As both Yeh (1978) and Fleming (1979) have pointed out, teachers report that they rely the most on observational data gathered during interactions

with students and their own intuitive hunches when making their instructional decisions. However, they also seemed to realize that these data were viewed as more subjective than standardized test results. Given this dilemma, it would seem that measurement educators need to focus on teaching practitioners ways to make these data more reliable and valid so that teachers can increase their confidence in using them.

7. Given the measurement needs described in 1-6, teachers reported they thought they could learn more about measurement concepts if the concepts were presented in such a way that (1) they were related to specific instructional problems encountered in the classroom, (2) they referred to specific children in the teacher's classroom, (3) they took into account the context of the teacher's classroom, and (4) they were organized around decisions made throughout the entire school year.

Professional Development

1. A professional development experience that allows for the interaction among teachers is a crucial

REPORTS AVAILABLE

Back issues of *Measurement in Education* are available for $2.50 each per single issue, and for 75¢ each in quantities of 25 or more for a single issue.

Vol. 1, No. 1 *Helping Teachers Use Tests* by Robert L. Thorndike
No. 2 *Interpreting Achievement Profiles — Uses and Warnings* by Eric F. Gardner
No. 3 *Mastery Learning and Mastery Testing* by Samuel T. Mayo
No. 4 *On Reporting Test Results to Community Groups* by Alden W. Badal & Edwin P. Larsen

Vol. 2, No. 1 *National Assessment Says* by Frank B. Womer
No. 2 *The PLAN System for Individualizing Education* by John C. Flanagan
No. 3 *Measurement Aspects of Performance Contracting* by Richard E. Schutz
No. 4 *The History of Grading Practices* by Louise Witmer Cureton

Vol. 3, No. 1 *Using Your Achievement Test Score Reports* by Edwin Gary Joselyn & Jack C. Merwin
No. 2 *An Item Analysis Service for Teachers* by Willard G. Warrington
No. 3 *On the Reliability of Ratings of Essay Examinations* by William E. Coffman
No. 4 *Criterion-Referenced Testing in the Classroom* by Peter W. Airasian and George F. Madaus

Vol. 4, No. 1 *Goals and Objectives in Planning and Evaluation: A Second Generation* by Victor W. Doherty and Walter E. Hathaway
No. 2 *Career Maturity* by John O. Crites
No. 3 *Assessing Educational Achievement in the Affective Domain* by Ralph W. Tyler
No. 4 *The National Test-Equating Study in Reading (The Anchor Test Study)* by Richard M. Jaeger

Vol. 5, No. 1 *The Tangled Web* by Fred F. Harcleroad
No. 2 *A Moratorium? What Kind?* by William E. Coffman
No. 3 *Evaluators, Educators, and the Publics: A Detente?* by William A. Mehrens
No. 4 *Shall We Get Rid of Grades?* by Robert L. Ebel

Vol. 6, No. 1 *On Evaluating a Project: Some Practical Suggestions* by John W. Wick
No. 2 *Measuring Reading Achievement: A Case for Criterion — Referenced Testing and Accountability* by S. Jay Samuels and Glenace E. Edwall
No. 3 *Security in a Citywide Testing Program* by Anthony J. Polemeni
No. 4 *Dear Mama: Why don't they love me anymore?* by Thomas J. Fitzgibbon

Vol. 7, No. 1 *The Discrepancy Evaluation Model (Part I)* by Andres Steinmetz
No. 2 *The Discrepancy Evaluation Model (Part II)* by Andres Steinmetz
No. 3 *What Graduate and Professional School Students Think About Admission Tests* by Leonard L. Baird
No. 4 *The Paradox of Education Testing* by Robert L. Ebel

Vol. 8, No. 1 *FACE UP: A Framework for Assessing Career Educators Utilization Programs* by Bruce W. Tuckman
No. 2 *Grade Equivalent Scores* by Gary Echternacht and *If Not Grade Equivalent Scores — Then What?* by Jeanne M. Plas
No. 3 *Criticisms of Testing: How Mean is the Median?* by Henry S. Dyer
No. 4 *The Superintendent and Testing: Implications for the Curriculum* by Herbert C. Rudman

Vol. 9, No. 1 *A Consumer's Guide to Criterion — Referenced Test Item Statistics* by Ronald C. Berk
No. 2 *Perspective on Intelligence Testing*, by Roger T. Lennon
No. 3 *Bias*, by Eric F. Gardner
No. 4 *The Use of Tests in Admissions to Higher Education* by Mary Fruen

Vol. 10,No. 1 *Some Comments on Criterion-Referenced and Norm-Referenced Achievement Tests* by William A. Mehrens and Robert L. Ebel
No. 2 *Does the Rasch Model Really Work? A Discussion for Practitioners* by R. Robert Rentz and Charlotte C. Rentz
No. 3 *A Perspective On The Uses and Misuses of Standardized Achievement Tests* by Peter W. Airasian
No. 4 *The Practical Side of Criterion-Referenced Test Development* W. James Popham and Elaine Lindheim

Vol. 11,No. 1 *The Courts and Your Minimum Competency Testing Program — A Guide to Survival* by Thomas H. Fisher
No. 2 *The Anatomy of a Scholastic Aptitude Test* by Roger T. Lennon

Vol. 12,No. 1 *Attitudes Toward Standardized Tests: Students, Teachers and Measurement Specialists* by Frank P. Stetz and Michael D. Beck
No. 2 *Issues in Testing for Competency* by Robert L. Ebel and Samuel A. Livingston
No. 3 *Protesting the Detesting of PRO Testing* by Jason Millman

element for long term success. As one teacher reported, "A real plus of this project was being able for four of us to meet for extended periods, and work on something useful to us."

2. The opportunity to have teachers from various grade levels work together increases the chance for a long term interaction after the project has ended.

3. Professional development must meet a need that teachers have identified. It is clear from the measurement literature that teachers, administrators and measurement experts do not share the same perceptions about what measurement knowledge teachers have a need (Stetz and Beck, 1981 and Yeh, 1978). While no one group probably has "the right answer", it is clear that unless teachers have ownership of the professional development experience, change will not occur. (Nicholson *et al.*, 1976 and Cruickshank *et al.*, 1979).

4. Administrative support of the professional development activities is crucial for success. The teachers must be aware of this support.

5. Consideration needs to be given to how educators can best measure changes in levels of teacher's knowledge as a result of professional development experiences. How specific an impact would it be realistic to expect?

The purpose of offering these reflections is to stimulate discussion and research/development activities related to these issues. It seems imperative that educational measurement experts continue to tackle the problem of increasing teacher's use of assessment data in instructional decision-making. It is time for measurement experts and school practitioners to jointly address this problem in a mutually beneficial manner.

References

Bradley, R. The teaching of testing. *Measurement News*, 1978, *21*, 4.

Cruickshank, D., Lorish, C., and Thompson, L. What we think we know about inservice education. *Journal of Teacher Education*, 1979, *30*, 27-32.

Development of Materials for Integrating Assessment with Instruction. (RFP-NIE-79-0038). Washington, D.C.: National Institute of Education, July, 1979.

Fleming, M. Classroom measurement needs in 1980's. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, 1979.

Lerner, B. "The War on Testing: David, Goliath and Gallup." *The Public Interest*, 1980, *60*, 119-147.

Mehrens, W., and Lehmann, I. *Measurement and Evaluation in Education and Psychology* (2nd ed.). New York: Holt, Rinehart and Winston, 1978.

Nicholson, A., Joyce, B., Parker, D., and Waterman, F. *The Literature on Inservice Teacher Education*, Syracuse, New York: The National Dissemination Center of Syracuse University, 1976.

Olejnik, S.F. Standardized achievement testing programs viewed from the perspective of the non-measurement specialist. Paper presented at the annual meeting of the National Conference on Measurement in Education, San Francisco, 1979.

Rudman, H.C., Kelly, J., Wanous, D., Mehrens, W., Clark, C., and Porter, A. *Integrating Assessment with Instruction: A Review.* (Research Series No. 75). East Lansing, MI: The Institute for Research on Teaching, Michigan State University, 1980.

Rudman, H.C., Mehrens, W.A., Wanous, D.S., Porter, A.C. and Clark, C.M. *Integrating Assessment with Instruction, 1979-1981: A Report.* Part I, Part II East Lansing, MI: Institute for Research on Teaching, Michigan State University, November, 1981.

Shanker, A. "The Nonsense of Attaching Education Tests." *The Washington Post.* October 19, 1980.

Stetz, F., and Beck, M. Attitudes toward standardized tests: Students, teachers and measurement specialists. *Measurement in Education*, 1981, *12*, 1-11.

Yeh, J.P. *Test Use in Schools.* Washington, D.C.: U.S. Department of Health, Education and Welfare, National Institute of Education, 1978.

*The authors wish to acknowledge the National Institute of Education, the Institute for Research on Teaching, and the Department of Defense Dependent Schools for their support of this project on which this paper draws (NIE Contract Number 400-79-0067). The project team included Herbert C. Rudman (Project Director), Christopher M. Clark and Andrew C. Porter as well as the authors of this paper. The ideas and opinions expressed in this paper are solely the responsibility of the authors.

A Few Closing Comments

What makes membership in NCME so enjoyable is the mix one encounters between measurement theory and its application to the practice of education. These pages, since Volume 1, number 1 have reflected this mix. *Measurement in Education* ceases to exist with the publication of this issue. It now takes a new form, a new name, and a new editor. The form will be a journal not unlike *Education Researcher;* the title of the new journal will be *Educational Measurement: Issues and Practice;* the editor will be an old friend and colleague, Dr. Frank B. Womer of the University of Michigan. The form, title, and editor will be new, but the mix of measurement theory and practice will remain. I wish for Dr. Womer the best of experiences and that those experiences be as fruitful and fulfilling for him as they have been for me.

I must thank those listed on the masthead who served so faithfully as the editorial advisory board these past three years. They were thoughtful, they were careful, and never once forgot that prospective authors were sensitive and concerned about the work they submitted. When articles were judged to be inappropriate for *ME* they said so forthrightly but tactfully and always with a word of encouragement. I valued their judgements; they made *ME* maintain its practical bent, and its quality.

I cannot close this issue without a word of appreciation to Ms. Timber Sue Weaver and Mrs. Melinda Russell of the Colorcraft Corporation who helped improve the design of *Measurement in Education* and who shepherded each manuscript through production. They are two friends I shall miss very much.

Finally allow me to thank you, the readers, for your support and willingness to submit manuscripts when invited and when you felt you had something important to share with others. You made my job as editor, "a piece of cake". **HCR**